

What is claimed is:

1 1. A method of determining a placement of services of a distributed application
2 onto nodes of a distributed resource infrastructure comprising the steps of:
3 a. establishing a placement indicator for a specific service;
4 b. forming communication constraints between node pairs which ensure that
5 a sum of transport demands between a particular node pair does not exceed a
6 transport capacity between the particular node pair, each term of the sum
7 comprising a product of a first placement variable, a second placement
8 variable, and the transport demand between the services associated with the
9 first and second placement variables;
10 c. forming an objective; and
11 d. employing a local search solution to solve an integer program comprising
12 the placement indicator, the communication constraints, and the objective to
13 determine the placement of the services onto the nodes.

1 2. The method of claim 1 wherein the placement indicator comprises a pre-
2 defined placement.

1 3. The method of claim 2 wherein the pre-defined placement comprises placing
2 the specific service onto a specific node.

1 4. The method of claim 2 wherein the pre-defined placement comprises not
2 placing the specific service onto a specific node.

1 5. The method of claim 1 wherein the placement indicator comprises a neutral
2 indication of whether the specific service is to be placed onto a specific node.

1 6. A method of determining a placement of services of a distributed application
2 onto nodes of a distributed resource infrastructure comprising the steps of:
3 a. establishing an application model of the services comprising transport
4 demands between the services;
5 b. establishing an infrastructure model of the nodes comprising transport
6 capacities between the nodes;

7 c. establishing a placement model comprising placement indicators for the
8 services;

9 d. forming an integer program that comprises:

10 i. a set of placement variables for a combination of the services and the
11 nodes, each of the placement variables indicating whether a particular
12 service is located on a particular node;

13 ii. communication constraints between node pairs which ensure that a
14 sum of the transport demands between a particular node pair does not
15 exceed the transport capacity between the particular node pair, each term
16 of the sum comprising a product of a first placement variable, a second
17 placement variable, and the transport demand between the services
18 associated with the first and second placement variables;

19 iii. placement constraints for the services which ensure that the services
20 are placed onto the nodes in accord with the placement indicators; and

21 iv. an objective; and

22 e. employing a local search solution to solve the integer program which
23 determines the placement of the services onto the nodes.

1 7. The method of claim 6 wherein a particular placement indicator comprises an
2 indication that a specific service is to be placed onto a specific node.

1 8. The method of claim 6 wherein a particular placement indicator comprises an
2 indication that a specific service is not to be placed onto a specific node.

1 9. The method of claim 6 wherein a particular placement indicator comprises a
2 neutral indication of whether a specific service is to be placed onto a specific
3 node.

1 10. The method of claim 9 wherein a default for the placement indicators
2 comprises the neutral indication.

1 11. A method of determining a placement of services of a distributed application
2 onto nodes of a distributed resource infrastructure comprising the steps of:
3 a. establishing an application model of the services that comprises processing

4 demands for the services, storage demands for the services, and transport
5 demands between the services;

6 b. establishing an infrastructure model of the nodes that comprises processing
7 capacities for the nodes, storage capacities for the nodes, and transport
8 capacities between the nodes;

9 c. establishing a placement model comprising placement indicators for the
10 services;

11 d. forming an integer program that comprises:

12 i. a set of placement variables for a combination of the services and the
13 nodes, each of the placement variables indicating whether a particular
14 service is located on a particular node;

15 ii. processing constraints which ensure that a sum of the processing
16 demands for each of the nodes does not exceed the processing capacity for
17 the node;

18 iii. storage constraints which ensure that a sum of the storage demands for
19 each of the nodes does not exceed the storage capacity for the node;

20 iv. first placement constraints which ensure that each of the services is
21 placed on one and only one node;

22 v. second placement constraints which ensure that the services are placed
23 onto the nodes in accord with the placement indicators;

24 vi. communication constraints between node pairs which ensure that a
25 sum of the transport demands between a particular node pair does not
26 exceed the transport capacity between the particular node pair, each term
27 of the sum comprising a product of a first placement variable, a second
28 placement variable, and the transport demand between the services
29 associated with the first and second placement variables; and

30 vii. an objective of minimizing communication traffic between the nodes
31 and balancing processing loads on the nodes; and

32 e. employing a local search solution to solve the integer program which
33 determines the placement of the services onto the nodes.

1 12. A computer readable memory comprising computer code for directing a
2 computer to make a determination of a placement of services of a distributed
3 application onto nodes of a distributed resource infrastructure, the determination

4 of the placement of the services onto the nodes comprising the steps of:
5 a. establishing a placement indicator for a specific service;
6 b. forming communication constraints between node pairs which ensure that
7 a sum of transport demands between a particular node pair does not exceed a
8 transport capacity between the particular node pair, each term of the sum
9 comprising a product of a first placement variable, a second placement
10 variable, and the transport demand between the services associated with the
11 first and second placement variables;
12 c. forming an objective; and
13 d. employing a local search solution to solve an integer program comprising
14 the placement indicator, the communication constraints, and the objective to
15 determine the placement of the services onto the nodes.

1 13. The computer readable memory of claim 12 wherein the placement indicator
2 comprises a pre-defined placement.

1 14. The computer readable memory of claim 13 wherein the pre-defined
2 placement comprises placing the specific service onto a specific node.

1 15. The computer readable memory of claim 13 wherein the pre-defined
2 placement comprises not placing the specific service onto a specific node.

1 16. The computer readable memory of claim 12 wherein the placement indicator
2 comprises a neutral indication of whether the specific service is to be placed onto
3 a specific node.

1 17. A computer readable memory comprising computer code for directing a
2 computer to make a determination of a placement of services of a distributed
3 application onto nodes of a distributed resource infrastructure, the determination
4 of the placement of the services onto the nodes comprising the steps of:
5 a. establishing an application model of the services comprising transport
6 demands between the services;
7 b. establishing an infrastructure model of the nodes comprising transport
8 capacities between the nodes;

9 c. establishing a placement model comprising placement indicators for the
10 services;

11 d. forming an integer program that comprises:

12 i. a set of placement variables for a combination of the services and the
13 nodes, each of the placement variables indicating whether a particular
14 service is located on a particular node;

15 ii. communication constraints between node pairs which ensure that a
16 sum of the transport demands between a particular node pair does not
17 exceed the transport capacity between the particular node pair, each term
18 of the sum comprising a product of a first placement variable, a second
19 placement variable, and the transport demand between the services
20 associated with the first and second placement variables;

21 iii. placement constraints for the services which ensure that the services
22 are placed onto the nodes in accord with the placement indicators; and

23 iv. an objective; and

24 e. employing a local search solution to solve the integer program which
25 determines the placement of the services onto the nodes.

1 18. The computer readable memory of claim 17 wherein a particular placement
2 indicator comprises an indication that a specific service is to be placed onto a
3 specific node.

1 19. The computer readable memory of claim 17 wherein a particular placement
2 indicator comprises an indication that a specific service is not to be placed onto a
3 specific node.

1 20. The computer readable memory of claim 17 wherein a particular placement
2 indicator comprises a neutral indication of whether a specific service is to be
3 placed onto a specific node.

1 21. The computer readable memory of claim 20 wherein a default for the
2 placement indicators comprises the neutral indication.

1 22. The computer readable memory of claim 20 wherein a matrix is specified

2 which expresses constraints or preferences for identifying a placement of services
3 onto nodes.

1 23. A computer readable memory comprising computer code for directing a
2 computer to make a determination of a placement of services of a distributed
3 application onto nodes of a distributed resource infrastructure, the determination of
4 the placement of the services onto the nodes comprising the steps of:

- 5 a. establishing an application model of the services that comprises processing
6 demands for the services, storage demands for the services, and transport
7 demands between the services;
- 8 b. establishing an infrastructure model of the nodes that comprises processing
9 capacities for the nodes, storage capacities for the nodes, and transport
10 capacities between the nodes;
- 11 c. establishing a placement model comprising placement indicators for the
12 services;
- 13 d. forming an integer program that comprises:
 - 14 i. a set of placement variables for a combination of the services and the
15 nodes, each of the placement variables indicating whether a particular
16 service is located on a particular node;
 - 17 ii. processing constraints which ensure that a sum of the processing
18 demands for each of the nodes does not exceed the processing capacity for
19 the node;
 - 20 iii. storage constraints which ensure that a sum of the storage demands for
21 each of the nodes does not exceed the storage capacity for the node;
 - 22 iv. first placement constraints which ensure that each of the services is
23 placed on one and only one node;
 - 24 v. second placement constraints which ensure that the services are placed
25 onto the nodes in accord with the placement indicators;
 - 26 vi. communication constraints between node pairs which ensure that a
27 sum of the transport demands between a particular node pair does not
28 exceed the transport capacity between the particular node pair, each term
29 of the sum comprising a product of a first placement variable, a second
30 placement variable, and the transport demand between the services
31 associated with the first and second placement variables; and

32 vii. an objective of minimizing communication traffic between the nodes
33 and balancing processing loads on the nodes; and
34 e. employing a local search solution to solve the integer program which
35 determines the placement of the services onto the nodes.